A MODEL-DRIVEN BIG DATA ARCHITECTURE FOR PLANETARY DATA ARCHIVES AND RESEARCH

https://pds.nasa.gov/

Daniel J. Crichton[1], J. Steven Hughes[1], Sean Hardman[1], Emily Law[1], Thomas C. Stein[2], Reta Beebe[3]

[1]Jet Propulsion Laboratory, California Institute of Technology [2]Washington University, [3]New Mexico State University

November 2017

Planetary Data System

- Purpose: To collect, archive and make accessible digital data and documentation produced from NASA's exploration of the solar system from the 1960s to the present.

- Infrastructure: A highly distributed software infrastructure with planetary science data repositories implemented at major government labs and academic institutions
  - System driven by a well defined planetary science information model
  - Over 1 PB of data
  - Movement towards international interoperability through IPDA
  - Distributed federation of US nodes and international archives

- Being realized through PDS4
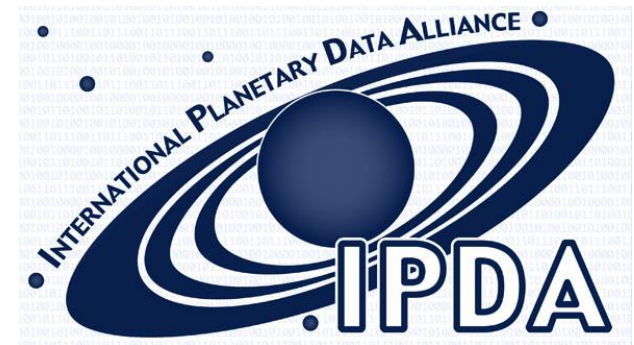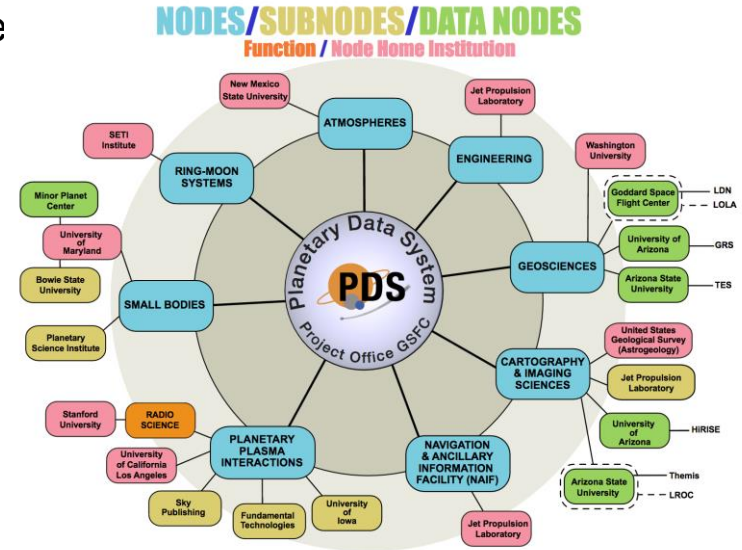
Key Drivers Impacting PDS

- More Data
- More Complexity (instruments, data)
- More Producer Interfaces
- Greater User Expectations
- Archive vs. Usability

- Funding Constraints
- Creating a system from the federation
- Internationalization
- Increasing IT security threats
- New Technology Opportunities

"Support the ongoing effort to evolve the Planetary Data System from an archiving facility to an effective online resource for the NASA and international communities." -- Planetary Science Decadal Survey, NRC, 2013-2022

Growth of Planetary Data Archived from U.S. Solar System Research

U.S. Planetary Data Archives (TBs)
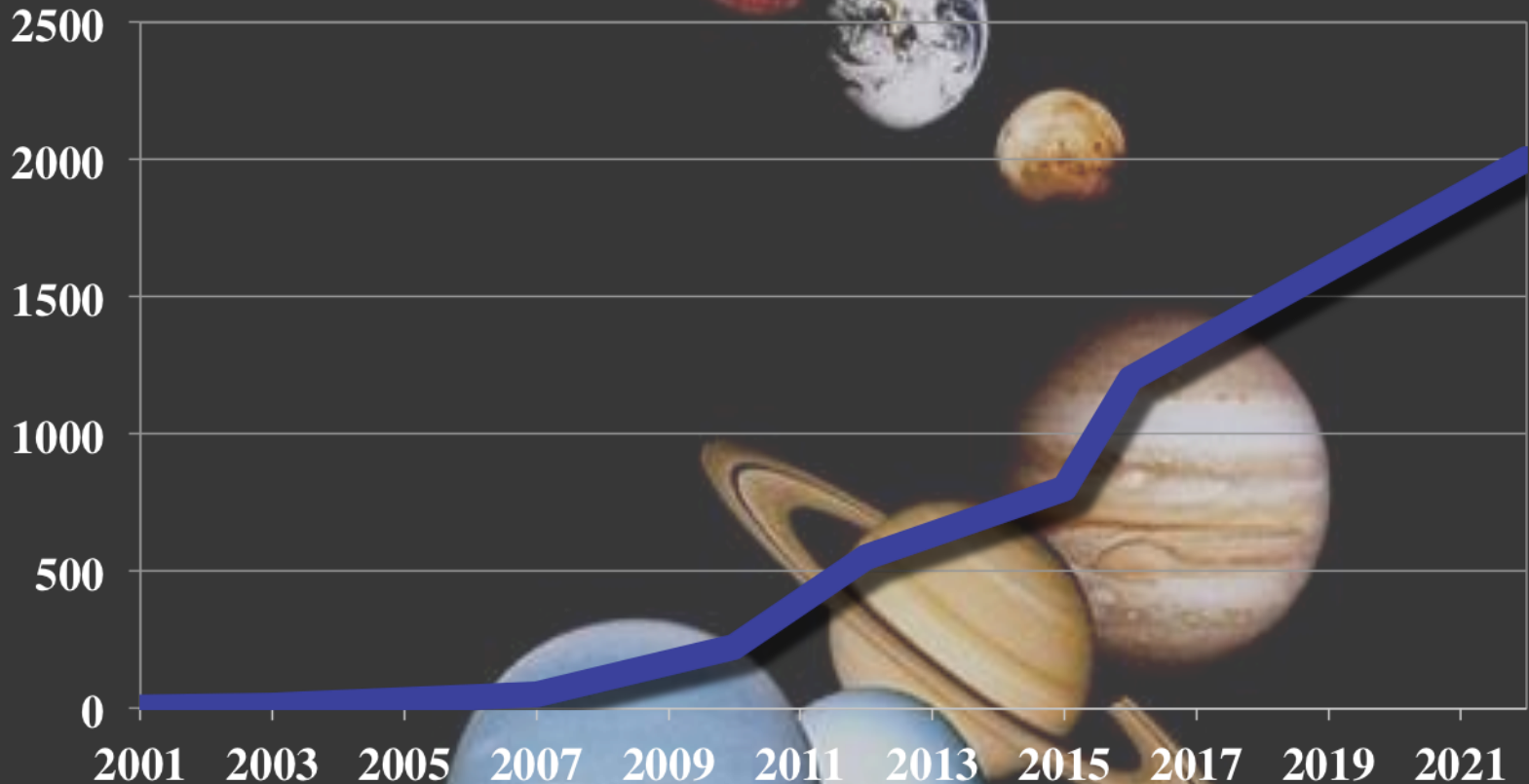
Diversity of the PDS

Type of Data	Distinct Products
Data Sets	2151
Instrument Hosts	199
Instruments	625
Targets	4231
Missions/Investigations	71
Volumes	5847

- Total volume is currently ~1.3PB
- Over 40M data products
- Some missions have few instruments but many data products, e.g., LADEE
- Variety of data is the challenge

PDS4

- An international, information model-driven data architecture for distributed planetary data archives

- An explicit information model
  - Explicitly describe the diversity of planetary data
  - Drive the definition of data to enable management, search and analytics across PDS and IPDA

- Distributed software services architecture
  - Services both within PDS and at international partners
  - Consistent protocols for access to the data and services
  - A distributed registry and search infrastructure
  - Tools that are built on top of the PDS4 information model

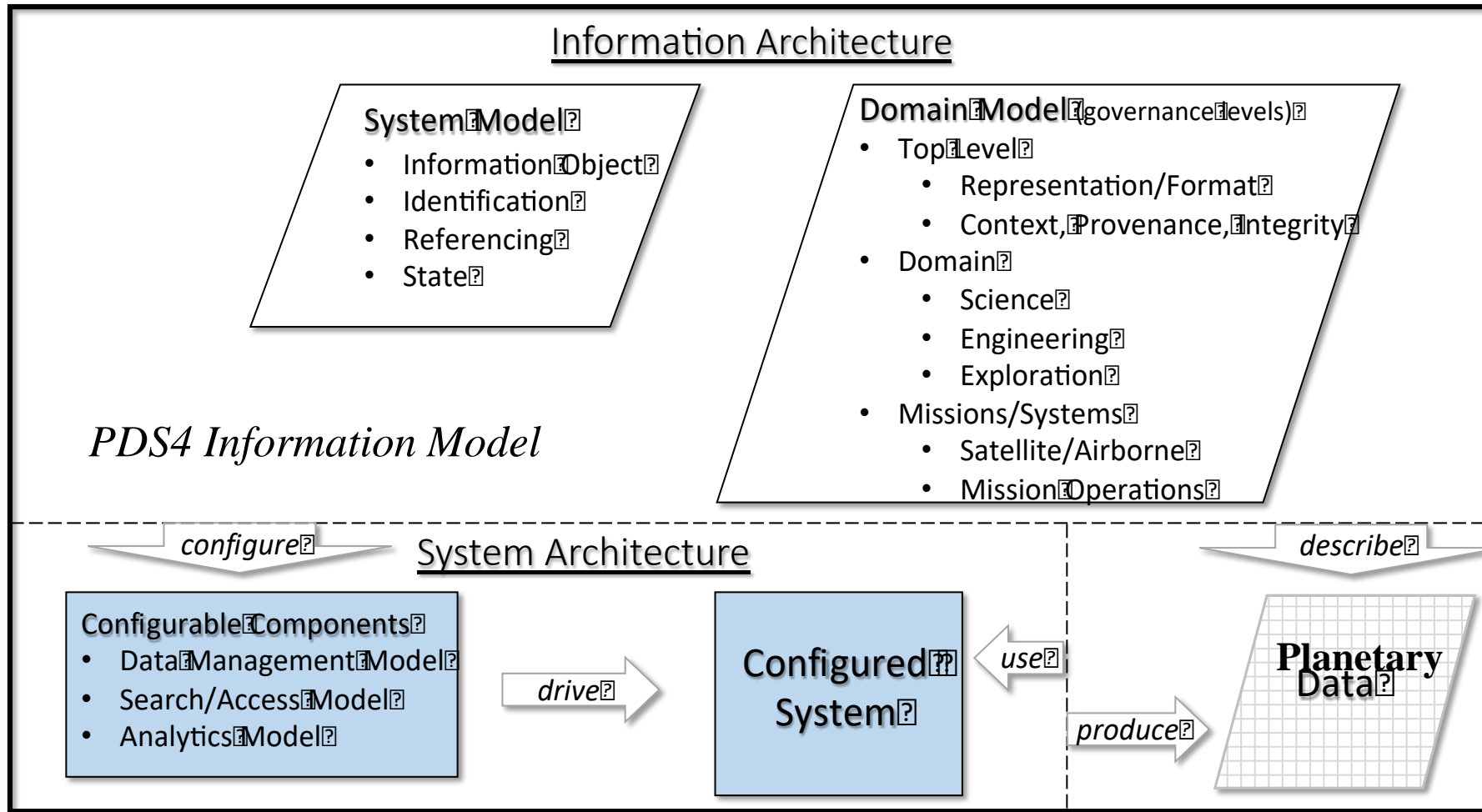PDS4 is co-developed with the community

PDS4 Information Model

- PDS4 Information Model plays a key role in defining the data and its relationships
  - Defines explicit relationships between major entities of the PDS
  - Establishes an overarching governance model for PDS data
  - Handles the diversity of different disciplines
  - New instruments, observation types and data can be accommodated
  - PDS metadat are tied to the model to increase consistency

- Changes managed through a change control board with members selected from PDS and International Community.

Model-Driven Information Architecture

- Allows data definitions to be exported from a modeling environment and into software
    - Is programming language agnostic
    - Allows the software and the definition of the data to evolve independently
    - Allows definitions of the data to be updated without changing software

- Drives an explicit data strategy for an organization

- Allows for the development of core data system libraries that work with the model

- Increases interoperability

PDS4: A Model-Driven Strategy
Information System Architecture

Information Architecture

System Model
- Information Object
- Identification
- Referencing
- State

Domain Model (governance levels)
- Top Level
  - Representation/Format
  - Context, Provenance, Integrity
- Domain
  - Science
  - Engineering
  - Exploration
- Missions/Systems
  - Satellite/Airborne
  - Mission Operations

PDS4 Information Model

configure

System Architecture

describe

Configurable Components
- Data Management Model
- Search/Access Model
- Analytics Model

drive

Configured System

use

Planetary Data

produce

9

Core System Builds

- PDS4 uses system builds to bring together the software and the information model
  - Established very early in the project to organize releases
  - Provides a predictable structure to bring the teams together
  - Provides incremental functionality relative to budget constraints
  - *Support co-development with the community*

- Each build provides a full lifecycle to capture, CM, integrate, test and deploy the release

- V1.9 released in November 2017

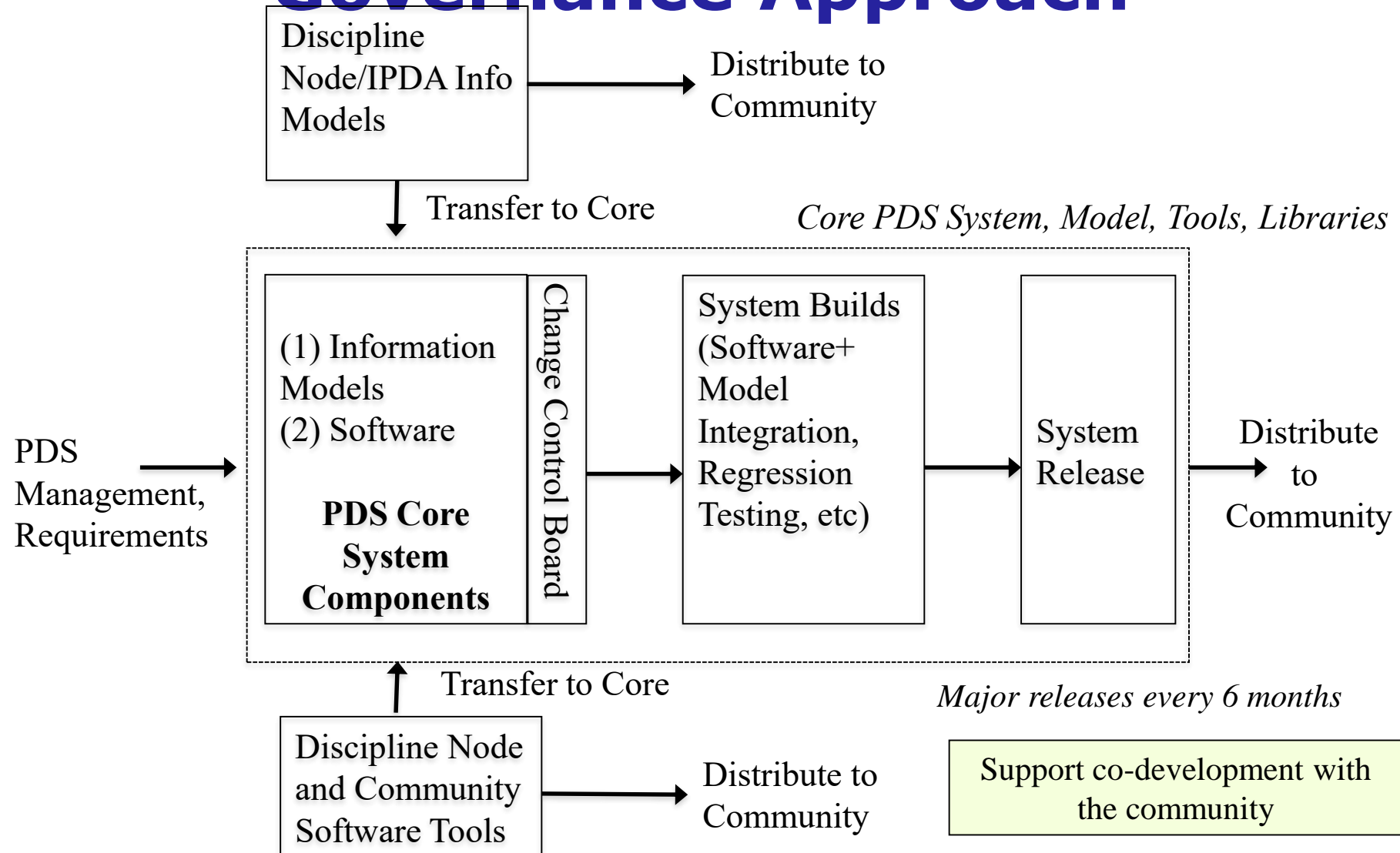# Core Software and Standards Deliverables

Software System

- Registry Service
- Harvest Tool
- Validate Tool
- Security Service
- Report Service
- Search Service
- Transform Tool
- Catalog Tool
- Tools for different types of registries
- PDS4 Libraries
- Upgraded portal search and page views to support PDS4

Data Standards*

- Information Model
- XML Schemas
- Data Dictionary
- Concepts Document
- Standards Reference
- Data Providers Handbook
- PDS4 Example Products

Posted to https://pds.nasa.gov/pds4

System Builds and Governance Approach

Discipline Node/IPDA Info Models

→ Distribute to Community

Transfer to Core

Core PDS System, Model, Tools, Libraries

PDS Management, Requirements →

(1) Information Models
(2) Software

PDS Core System Components

Change Control Board

System Builds (Software+ Model Integration, Regression Testing, etc)

System Release

→ Distribute to Community

Transfer to Core

Major releases every 6 months

Discipline Node and Community Software Tools

→ Distribute to Community

Support co-development with the community

Planetary Tools Registry

https://pds.nasa.gov/tools/tool-registry

International Planetary Data Alliance

In 2005, it was recognized that the international agencies should collaborate on coordinating data standards and sharing of planetary science data

This led to the formation of the International Planetary Data Alliance (IPDA) in 2006

http://planetarydata.org

The Mission of the IPDA...

- "Facilitate global access to, and exchange of, high quality scientific data products managed across international boundaries"

- Support construction of compatible archives

- Support sharing of tools and software services

IPDA Members

ASI		ISRO	
CNSA		JAXA	
CSA		NASA	
CNES		Russia/IKI	
DLR		United Arab Emirates	
ESA		UK Space Agency	

International Collaboration on PDS4 Through IPDA
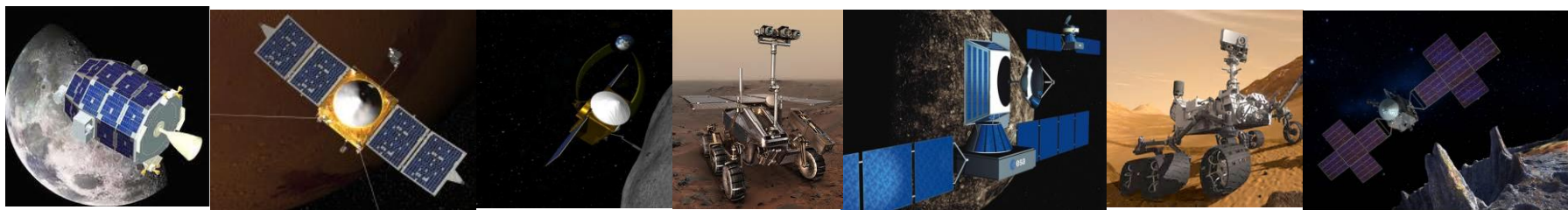
LADEE (NASA)   MAVEN (NASA)   Osiris-REx (NASA)   ExoMars (ESA/Russia)   BepiColombo (ESA/JAXA)   Mars 2020 (NASA)   Psyche (NASA)
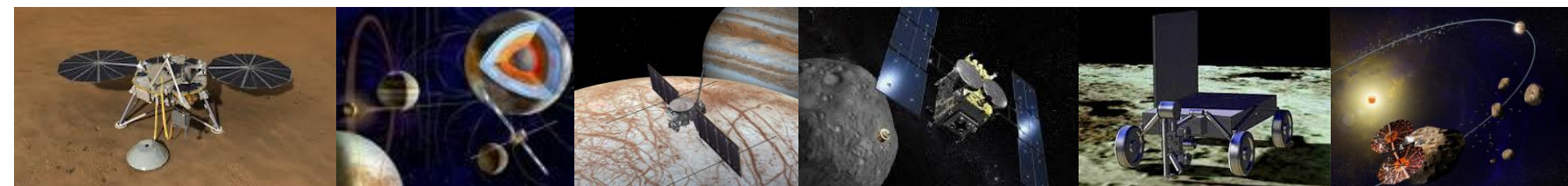
InSight (NASA)   JUICE (ESA)   Europa (NASA)   Hyabussa-2 (JAXA)   Chandrayaan-2 (ISRO)   Lucy NASA

Endorsed by the **International Planetary Data Alliance** in July 2012 –
https://planetarydata.org/documents/steering-committee/ipda-endorsements-recommendations-and-actions

International Registration, Search, and Access

planetarydata.org

pds.nasa.gov

Other Search Portals / Applications

Search Service supports the PDS and PDAP protocols enabling development of other portals and applications on this infrastructure.

PDS Search Service

PDS Search Service

Common PDS and IPDA Protocols

Indexed Metadata

Indexed Metadata

Registered Objects:
• Websites
• Data Sets
• Investigations, Instruments, etc.
• Tools and Services

PDS Registry Service

Data Set Metadata

PDS Registry Service

PDAP-Based Interfaces

PDS Registry Interface

Planetary Science Archive

Other International Archives

PDS Discipline Node Archives

Search results include mission support pages and other more specific search interfaces.

18

Future: An International Platform for Planetary Data Archiving, Management, and Research
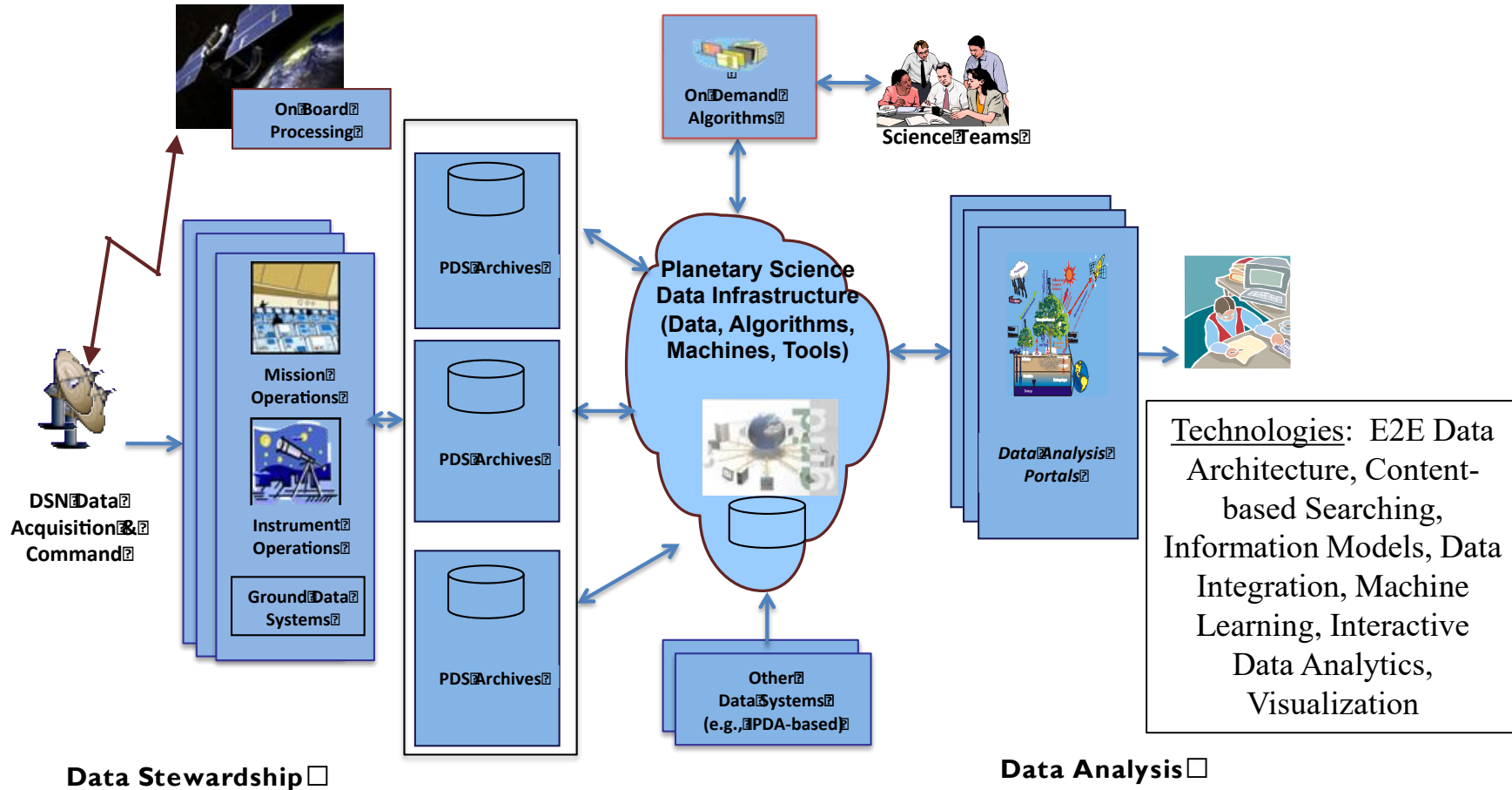
On Board Processing

On Demand Algorithms

Science Teams

DSN Data Acquisition & Command

Mission Operations

Instrument Operations

Ground Data Systems

PDS Archives

PDS Archives

PDS Archives

Planetary Science Data Infrastructure (Data, Algorithms, Machines, Tools)

Other Data Systems (e.g., IPDA-based)

Data Analysis Portals

Technologies: E2E Data Architecture, Content-based Searching, Information Models, Data Integration, Machine Learning, Interactive Data Analytics, Visualization

Data Stewardship

Data Analysis

"*Support the ongoing effort to evolve the Planetary Data System from an archiving facility to an effective online resource for the NASA and international communities.*" -- Planetary Science Decadal Survey, NRC, 2013-2022

1ˢᵗ Planetary Science Informatics & Data Analytics Conference

Data Management, Cloud Computing, Data Modeling, Information Systems, Information Processing, Machine Learning, Visualization, and more…

Summary

- PDS has successfully developed PDS4 with the international community
  - *All major space agencies collecting planetary science data committed to using PDS4 through the IPDA*

- Provides a novel, model-driven approach

- Good progress towards international interoperability and data sharing

- Increasing opportunities to enable data-driven approaches
  - *See upcoming Planetary Data Analytics Workshop in April 2018*